UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

DTIC FILE COPY

(1)

AD-A196 291

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFIT/CI/NR 88- 126	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A DETECTION THEORY ANALYSIS OF VISUAL DISPLAY PERFORMANCE		5. TYPE OF REPORT & PERIOD COVERED MS THESIS
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) THOMAS RICHARD MABRY		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS AFIT STUDENT AT: PURDUE UNIVERSITY		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS		12. REPORT DATE 1988
		13. NUMBER OF PAGES 53
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) AFIT/NR Wright-Patterson AFB OH 45433-6583		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

DISTRIBUTED UNLIMITED: APPROVED FOR PUBLIC RELEASE

DTIC ELECTE AUG 0 3 1988 S D

CD

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

SAME AS REPORT

18. SUPPLEMENTARY NOTES

Approved for Public Release: IAW AFR 190-1
LYNN E. WOLAVER Lynn Wolaver 21 July 88
Dean for Research and Professional Development
Air Force Institute of Technology
Wright-Patterson AFB OH 45433-6583

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ATTACHED

88 8

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT

Mabry, Thomas Richard M.S., Purdue University, December, 1987. A Detection Theory Analysis of Visual Display Performance. Major Professor: Robert D. Sorkin.

This study investigated how information is processed from graphic vs. alphanumeric multi-element visual displays using principles derived from the Theory of Signal Detection (TSD). A diagnostic decision task was used in an evaluation of four different display formats: (1) A numerical display composed of n two-digit numbers arranged in a linear horizontal format; (2) a similar numerical display in which the display elements were arranged in a square matrix array; (3) an analog gauge display composed of n vertical line gauges also organized in a square matrix array; and (4) a similar analog gauge display in which the display elements were arranged in a linear horizontal format. Performance was evaluated for 1, 2, 4, 9, and 16 element displays and over a range of display durations. Detection performance, as measured by d′, increased as the number of display elements was increased up to an asymptotic value that was dependent on display type, arrangement and display duration. Performance was best with analog display elements arranged in a horizontal line.

The relative influence of a particular spatial element and the total number of elements that influence a subjects's response appears to be highly dependent on display type and arrangement. When

processing information from numeric displays, only a few central elements influence the response. With analog displays, a greater number of display elements have a bearing on the subject's response. Detection theory methodology allows specification of important parameters of visual display processing and facilitates the comparison of different display types.

A DETECTION THEORY ANALYSIS OF VISUAL DISPLAY PERFORMANCE

A Thesis

Submitted to the Faculty

of

Purdue University

by

Thomas Richard Mabry

In Partial Fulfillment of the

Requirements for the Degree

of

Master of Science

December 1987

Dedicated to my family, Carol and Taylor. Your love, support and sacrifice made this possible. I could not have done it alone.

ACKNOWLEDGMENTS

I am truly indebted to Greg Elvers and Mike Snow for their friendship and invaluable help. I also want to thank Bob Hendrich for his assistance during data collection. Special thanks goes to Bob Sorkin, whose patience and advice guided me throughout this endeavor. This project was supported in part by grant AFOSR-84-0302 from the United States Air Force Office of Scientific Research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Mabry, Thomas Richard M.S., Purdue University, December, 1987. A Detection Theory Analysis of Visual Display Performance. Major Professor: Robert D. Sorkin.

This study investigated how information is processed from graphic vs. alphanumeric multi-element visual displays using principles derived from the Theory of Signal Detection (TSD). A diagnostic decision task was used in an evaluation of four different display formats: (1) A numerical display composed of n two-digit numbers arranged in a linear horizontal format; (2) a similar numerical display in which the display elements were arranged in a square matrix array, (3) an analog gauge display composed of n vertical line gauges also organized in a square matrix array; and (4) a similar analog gauge display in which the display elements were arranged in a linear horizontal format. Performance was evaluated for 1, 2, 4, 9, and 16 element displays and over a range of display durations. Detection performance, as measured by d', increased as the number of display elements was increased up to an asymptotic value that was dependent on display type, arrangement and display duration. Performance was best with analog display elements arranged in a horizontal line.

The relative influence of a particular spatial element and the total number of elements that influence a subjects's response appears to be highly dependent on display type and arrangement. When

processing information from numeric displays, only a few central elements influence the response. With analog displays, a greater number of display elements have a bearing on the subject's response. Detection theory methodology allows specification of important parameters of visual display processing and facilitates the comparison of different display types.

INTRODUCTION

The choice of display code and format is a critical step in the process of designing visual displays. The type, arrangement and number of displays can directly affect the performance of the intended user, yet there doesn't seem to be any general algorithm to guide these design decisions (Tullis, 1983). Today's sophisticated computer graphic capabilities highlight the need for such guidelines because hardware limitations and economics no longer dictate whether a display will be graphic or alphanumeric. A designer can now concentrate on optimizing human operator performance and not be concerned with exceeding the capabilities of the machinery that generates the visual displays. This study compares how information is processed from graphic vs. alphanumeric multi-element visual displays using principles derived from the Theory of Signal Detection (TSD). Two TSD based performance models will be discussed and evaluated. This study also investigates how subjects aggregate information from multi-element visual displays when display duration is too short to allow for full processing of all display information.

Previous Experiments

Empirical evidence directly comparing graphic vs. alphanumeric displays is sparse and the results are contradictory (Tullis,1981; DeSanctis, 1984). Some studies have shown display type had no effect

on performance. For example, Vicino and Ringel (1966) found that the timeliness and accuracy of decisions made in a tactical military scenario did not vary with graphics vs. alphanumeric presentation methods. In addition, Nawrocki (1973) found no significant advantages to either graphics or alphanumerics when subjects were required to remember information presented in a previous problem solving task.

In several instances, the superiority of alphanumerics has been demonstrated. Lincoln and Cahill (1965) found that subjects could more quickly detect "out of tolerance" conditions when system values were displayed with numbers rather than with analog meter gauges. Manipulating display duration did not effect results - performance on alphanumeric displays was superior across durations. A study by Grace (1966) also demonstrated the superiority of alphanumeric presentation of information in a task that tested information recall. Subjects presented with alphanumeric information on aircraft flight profiles scored higher on post-presentation tests than subjects given the same information in a graphic form. Remington and Williams (1986) demonstrated the superiority of alphanumeric characters in a visual search task. Subject reaction times were significantly faster for both present/ absent responses when the search target was an alphanumeric symbol instead of a graphic symbol. Error rates (saying a target was present when it was not and vice versa) were also lower when the target was an alphanumeric symbol. In addition, when new "markers" that represented additional information were added to the original target symbols, reaction times to all symbols increased but the increase was substantially greater for the graphic symbols.

There is a body of evidence, however, that has demonstrated the superiority of graphic displays. Crawford (1977) reviewed several studies that encompassed a wide range of skills and tasks. In each case, greater transfer of training occured when initial subject training utilized interactive computer based graphics rather than conventional training methods (i.e. textbooks and workbooks). Stock and Watson (1984) showed interpretation of financial indicators was facilitated by graphical presentation of information vs the conventional tabular presentation format. Tullis (1981) reports similar results when subjects were asked to interpret the results of a diagnostic test performed on a telephone system. The evaluation measured speed and accuracy of the subjects' interpretation of the test results presented in two graphic (black-and-white or color) or two alphanumeric (narrative or structured) formats. Accuracy did not vary across format but speed did. Response times for both graphic formats were consistently shorter than those of the narrative format and initially for the structured format as well. However, after considerable practice, RTs for the structured format did approach those for the graphics conditions. Tullis concluded that graphics certainly enhances performance for little practiced subjects and that this benefit is evident in very practiced subjects as well.

It is difficult to derive any general guidelines that might direct design decisions from such mixed empirical results, especially since the data come from such diverse tasks. Such guidelines have been proposed (see Tullis, 1983). However, the utility of such an approach

is questionable. Working within the framework of a theoretical model seems a much more promising approach.

Theoretical Approaches

Wickens (1984a) proposes the use of principles from Multiple Resource Theory (MRS) to optimize display design. The theoretical works of Kahneman (1973), Norman and Bobrow (1975) and Navon and Gopher (1979) serve as the underpinnings of MRS. The theory assumes that an inferred underlying commodity, of limited availability, called Resources, enables performance of a task. In a multi-task situation, each task must compete for resources, which may result in performance decrements. But unlike Kahneman (1973) who assumes that there is one central pool of resources with satellite structures, MRS supposes that humans possess several different capacities with resource properties (Wickens, 1984b). Therefore, tasks will interfere more if more of the resources required to accomplish each task are shared. Wickens (1984a) has defined these different resources along three dichotomous dimensions: 1) Stage (Encoding/Central Processing vs. Responding); 2) Code (Spatial vs. Verbal); and 3) Modality (Visual vs. Auditory). Multi-component displays should tap different resources in order to optimize operator performance. Since the present task requires a single decision to be made based on information presented only visually from multiple displays, the MRS approach is not applicable to the present situation.

MacGregor and Slovic (1986) have proposed use of the Brunswik (1956) lens model to determine optimal display format in a multiple cue situation. The model is based on three elements that are present in

any choice or decision situation: the stimulus cues, the correct response or answer, and the subject's response. Several aspects of performance such as stimulus diagnostic power, subject accuracy and achievement, subject consistency, and how well the subject's task strategy "matches" the optimal task strategy can be determined from regression equations and multiple correlation calculations between the basic elements (Dudycha and Naylor, 1966). MacGregor and Slovic (1986) demonstrated the viability of using a lens model approach to evaluate different graphic display types.

Sorkin and Weldon (unpublished) applied general principles from the Theory of Signal Detection (TSD) to the problem of evaluating the effects of different visual display formats on performance (Green and Swets, 1974). They selected a diagnostic decision task to compare the accuracy of detection performance on two different analog displays and one type of digital display.

Subjects viewed multi-element visual displays on a computer monitor. Either 1, 2, 4, 9, or 16 elements comprised each display. The elements were either two digit numbers (e.g. -0.5, 1.6) or meter-type analog gauges. Element arrangement was also manipulated. The elements were arranged in a horizontal line or in a square matrix array. After viewing each multi-element display for the defined observation interval, subjects had to report whether the displayed elements had been generated from one of two possible underlying normal distributions - the "signal" or the "noise" distribution. The display element values presented on each trial were generated according to statistical rules determined by whether a signal-plus-noise or noise-

alone condition was present. Each display element value was generated independently and all elements were generated from the same distribution on each trial. The "signal" distribution had a mean value of 0.85 and the "noise" distribution had a mean value of zero. Possible element values ranged from -2.0 to 2.0. The variance of each distribution was equal to unity. Consequently, the "signal" condition produced larger average values than the "noise" condition. The task was thus formally identical to a conventional signal detection task in which a subject must make a signal (yes) or non-signal (no) response.

TSD makes strong predictions about maximum or "ideal" performance for a single display element case. Performance for an ideal statistical observer (d'_{ideal}) is given by the difference between the means of the signal and noise distributions divided by their standard deviation (Green and Swets, 1974). So in the Sorkin and Weldon study:

$$d'_{ideal} = 0.85 / 1.0 = 0.85 \qquad [1]$$

How do subjects aggregate information from the multiple element displays? An appropriate statistic (one that is monotonic with the likelihood ratio basis of TSD) is the <u>sum</u> of the observations made on the individual elements. Since all display values are independently generated from either the signal or the noise distribution on each trial, the mean of this sum statistic would equal "n" times the mean of the distribution of each element sampled on that trial, where "n" is the number of elements sampled. Using the above example, this sum statistic would have a mean of 0.85n on the signal trials and a mean of zero on the noise trials. The standard deviation of the sum statistic

would be $n^{1/2}$. The difference between the means of the signal and noise distributions divided by their standard deviation equals

$$0.85n \,/\, n^{1/2} \qquad [2]$$

thus $d'_{(ideal)} = 0.85n^{1/2}$. So ideal performance in a multi-element display will increase with $n^{1/2}$ (assuming the ideal observer can combine the independent observation decisions with no information loss) (Green and Swets, 1974).

Several empirical tests of this performance prediction have been accomplished and results within the visual domain show general agreement with the $n^{1/2}$ prediction (see Swets, 1984). These empirical results have also led to the development of mathematical models of observer signal processing and decision behavior to account for deviations from ideal behavior. Swets (1984) conducted a thorough review of these experiments and of the various models developed for detection performance. Swets' (1984) review argued that TSD analysis was a useful approach to compare and contrast data from quite different experiments. It also showed there was general agreement with the $n^{1/2}$ performance prediction across a wide range of experimental data.

Based on those data, Sorkin and Weldon predicted that performance in their experiment would increase as a function of the number of display elements, but not necessarily at the square root of n rate for all experimental display formats and conditions. If the subject encountered some processing interference among the various display elements, then performance could be less than ideal. A non optimum combination rule applied to the measure made on each display element

would also result in less than ideal performance. The presence of some additional internal noise common to several display elements could cause performance to deviate from the ideal as well. In general, any fixed source of subject variability or internal noise would result in performance reaching an asymptotic level as n is increased.

The concept of internal noise is a central feature of the Partitioned Variance Model (Robinson, Grantham, and Berg, 1986; Berg, 1987; and Sorkin, Robinson, & Berg, 1987). It offers an explanation for human performance increases at less than the predicted $n^{1/2}$ rate. During the detection process, a human observer may suffer from additional sources of noise other than that associated with the signal (external noise) itself. For the Partitioned Variance Model, the performance equation is given by:

$$d'(n) = (M_{s+n} - M_n)/(V_{ext}/n + V_p/n + V_c)^{1/2} \qquad [3]$$

where internal noise is added at two stages: V_p is noise added before the decision statistic is formed ("peripheral" noise) and V_c is added after the decision is formed ("central" noise). The model assumes internal noise is added independently to each observation and a decision statistic is formed by averaging the n "noisy" observations. This is represented by the V_p term. The V_c term represents variability caused by decision criterion uncertainty, fluctuations of response bias, or memorial factors associated with the decision statistic. Central noise, V_c, is not dependent on the number of observations and is added to the decision statistic before this statistic is compared to the decision criterion (Berg, 1987).

Sorkin and Weldon also predicted that short duration displays, while allowing the aggregation of information from all the display elements, might result in reduced performance. They further hypothesized that if complete subject processing of display data was truncated by short duration stimulus presentations, then performance identical to that from a multi-element display having fewer elements would result. In other words, performance with short duration displays would increase as a function of the square root of the number of display elements and then reach an asymptotic value at some maximum value of n. Determining the asymptotic level of performance and the associated display duration for the various display formats and arrangements was a goal of the Sorkin and Weldon experiment.

Two general results obtained by Sorkin and Weldon provided justification for use of d' as a summary measure of subject performance. First, the Receiver Operating Characteristics (ROC) generated were consistent with TSD assumptions. The ROC curves appeared roughly symmetric, with approximately equal signal plus noise and noise variance on the decision statistic. Detectabilities implied by the rating data were consistent with values obtained in a yes/no task. Second, in many of the conditions tested, subject performance was very close to performance predicted for an ideal observer. Sorkin and Weldon concluded that this implied that in many conditions the subjects were able to utilize all the available information in the display when making the required detection decision.

Their results also showed a marked difference in performance between display formats. At the longest duration, 1020 ms, actual

performance closely approximated the ideal function for all display types. The greatest deviation occured when 9 or 16 elements were presented; the deviation from ideal was greatest in the numerical display, followed by the analog square array and by the linear analog. As display duration was decreased, performance continued to approximate the ideal function but asymptoted at different levels depending on display type. For example, performance on the numerical display at 255 ms seemed to asymptote at the N - 4 display element level. Except for the linear analog case which appeared independent of display duration, the number of display elements at asymptote decreased in an orderly fashion as display duration was decreased.

Sorkin and Weldon drew several conclusions from the data. First, depending on the display format and duration, subjects are able to aggregate information from as many as sixteen displays elements. Second, performance increased at close to the ideal rate up to different asymptotic levels, depending on the display format and display duration. Third, information was processed more efficiently from analog gauges. Finally, at short durations, information accumulation was highly dependent on display element arrangement.

The Present Experiment

One goal of the present experiment is to replicate the results obtained by Sorkin and Weldon. The same methodology will be employed with the following exceptions. The present experiment adds a linear numeric display format condition. This condition will allow for direct comparisons between linear analog and numeric formats. Also, Sorkin and Weldon used display values that ranged from -2.0 to 2.0. In order

to eliminate the possibility that subjects were simply using the signs of the display values and not evaluating the actual element values to make their detection decision, only positive display values will be used in the present study.

Another difference in the present experiment is the counterbalancing scheme employed. The Sorkin and Weldon experiment tested one type of display format at a time. Manipulations of display duration and number of display elements were counterbalanced across blocks within a fixed display type. So subjects completed work with one type of display format, e.g. linear analog gauges, before being exposed to a different format. The present experiment was partitioned by display duration. Manipulations of display format and number of display elements were counterbalanced across blocks within a fixed display duration. Once all conditions had been tested, display duration was then shortened. This counterbalancing scheme should minimize carryover practice effects caused by back-to-back presentations of similar stimuli.

One other manipulation of the Sorkin and Weldon paradigm was accomplished. Because of the simple makeup and homogeneous nature of display elements employed in this study, it was expected that the total visual angle covered by the display, rather than the density of the individual element displays, would be the main factor in determining how much information can be aggregated from the display within a given display duration. Consequently, two different display densities were be used. In one condition, the distance between elements was fixed at approximately 37.5 '. Thus, a larger number of display elements would

result in a greater total visual angle. In a second condition, the total visual angle was fixed at approximately 12.5 . In this case, a larger number of display elements would result in tighter spacing between the elements, down to approximately 21 ' for the n = 16 condition.

The main goal of the experiment was to determine how information is processed from a multi-element display when there was insufficient time to fully process all the information provided. The objective was to determine whether subjects continue to process information from all display elements, but at a degraded level, or whether subjects process information only from some asymptotic number of elements that is dependent on display type and display duration, and do not acquire information from the rest of the display. In addition, display design would be greatly facilitated if it could be determined whether information from some display elements weighs more heavily in the observers' decision than information from other elements in a display.

An analytic procedure developed by Robinson et al. (1986) was used to accomplish this objective (Berg, 1987). The procedure allows estimation of the relative contribution of each element in the stimulus display to the observer's decision statistic. This factor includes the effects of internal noise, attention, and response weight on the observer's decision. Sorkin, Robinson & Berg (1987) employed this method to estimate the relative contribution of each temporal position of a sequence of tones. The total number of tones in the sequence was varied from two to ten. The results clearly showed that the last tone in the sequence contributed the most to the decision. The first tones

in the sequence were the next most influential and the middle tones were least influential. This pattern was repeated for all possible tone lengths, n.

In the current experiment, distinct differences were predicted in the estimates of the relative contribution for each display position for different display types and arrangements. Because physiological data such as eye movement distances and patterns were not collected, no inferences about scanning patterns employed in searching the different types of displays could be made. However, estimates of the element contributions for the numeric linear display format were predicted to increase from left to right. This same pattern of results was also predicted to hold for the linear analog format, but to a lesser extent. For the square analog displays, it was hypothesized that the gauges at the center of the array near the fixation point would make the greatest contribution and that the contribution would decrease towards the periphery. This same pattern was hypothesized to characterize the square numeric format as well.

METHOD

This study has two goals. The first goal is to evaluate a TSD-based method for comparing different types of visual display codes and formats. Actual subject performance will be compared to TSD predictions. Parameter estimates for the Partitioned Variance Model will also be obtained. The second goal is to determine how subjects aggregate information from multi-element visual displays when display duration is too short to allow for full processing of all display information. Do subjects process information from all display elements at a reduced level, or is information processed from only a subset of the display elements while the others are ignored?

Two experiments, each comprised of two phases, were conducted. The second experiment, which was essentially a replication of the first, additionally assessed the effects of display density (spacing) on performance. The first phase of each experiment evaluated different visual display formats and codes within a TSD framework. The second phase of each experiment attempted to determine how subjects aggregate information at truncated display durations.

Experiment 1

Subjects

Four female undergraduates served as subjects. All subjects had normal or corrected-to-normal vision. Subjects were paid an hourly

wage plus a performance bonus for accuracy. Subjects were tested in two-hour sessions for approximately ten weeks.

Procedure

Phase One. The experimental task employed was a diagnostic decision task. On each trial, subjects viewed a multi-element visual display. The value for each display element was generated from one of two possible underlying normal distributions - the "Signal" or the "Noise" distribution. The subject's task was to determine which distribution had generated the display values on that trial. Subjects were instructed to try and use all of the information present during each trial. However, the experimenter gave no specific instructions on how to combine or integrate that information in order to make a decision. Subject instructions made it clear that for each trial all display values would come from the same distribution. Subjects clearly understood that about half of the trials would be Signal trials and that each display element value was generated independently. Subjects were also explicitly informed that the display values possible for both the Signal and Noise condition could span the range of allowable values (0.0 to 4.0 in 0.1 increments), but the average value for the Signal condition was 2.4 and the average value of the Noise condition was 1.6. No information concerning distribution standard deviation was provided.

Subjects responded by pressing one of four possible buttons labeled 'DEFINITELY NOISE," "PROBABLY NOISE," "PROBABLY SIGNAL," or "DEFINITELY SIGNAL." Subjects received performance feedback on the computer monitor at the end of each trial. Subjects were tested in pairs. Each pair of subjects received the same stimuli on each trial,

but subjects were unable to view the other subject's computer monitor, response selection, or feedback information.

A trial sequence consisted of the following events. A small fixation dot appeared in the center of the screen for 1000 ms. Five hundred milliseconds after fixation offset, the display(s) appeared for the appropriate duration. The entire screen was then blanked with a white masking screen for 1000 ms. Subjects had five seconds to respond from masking screen offset. Reaction time was not measured. Feedback as to whether the response was correct was then given at the bottom of the monitor screen. For each trial, the trial type (Signal or Noise), actual element values, and subject response were recorded. Subjects were given two minute rest breaks after each block, except between blocks four and five where a five minute rest break was given.

Apparatus

Stimuli were presented on IBM Personal Computer color monitors driven by a PC's Limited AT personal computer. Subjects sat 23 inches from the monitor in a dimly lit, sound isolated experimental chamber and recorded their decisions on handheld response boxes.

Monitors were adjusted for maximum contrast and a luminance of 27 footlamberts with the white blanking mask covering the screen. Because of the way in which the displays were generated, luminance varied for each different displays condition and for each different number of possible displays. However, in all cases the individual elements were completely legible at all display durations. Spragg and Rock (1952) reported that for high contrast displays, increases in luminance above a minimum level does not result in increments in dial reading

performance, so the luminance confound was not expected to contaminate results. The background illuminance level of the experimental chamber was 1.8 foot-candles.

Two display types were tested in the experiment: (1) two-digit numerical displays elements (digits) that were 0.25 inches high (visual angle = 0.625^0) by 0.375 inches wide (0.934^0); and (2) analog (meter type) display elements that were 0.375 inches wide (0.934^0), and 1.25 inches high (3.11^0). No numbers were displayed with the analog gauges. The reading for each gauge was represented by a bright white bar across the gauge. Tickmarks (one for every 0.5 increment) along the side of each gauge aided value determination. Displays presented were comprised of 1, 2, 4, 9, or 16 elements.

The display elements appeared in two formats: linear horizontal or square matrix array. In the square matrix array format, the gauges were centered on the display with an equal number of displays in the rows and columns. For both presentation formats the distance between the display elements was fixed, therefore, the greater the number of elements in the display, the greater the total visual angle covered by the display. Maximum visual angle measurements were: Horizontal = 12.86 degrees, Vertical = 14.62 degrees, for the sixteen element square matrix array condition; and Horizontal = 22.44 degrees for the sixteen element linear horizontal condition.

Display values were generated as follows. On each trial the computer randomly selected a deviate value for each display element from a normal distribution that had a mean of 1.6 and a standard deviation of 0.89. The computer then randomly determined whether a

signal would be present (probability of signal = 0.5) on that trial. If a Signal was chosen, then 0.8 was added to each generated value. Possible gauge values ranged from 0.0 to 4.0 in 0.1 increments. If a generated value exceeded these limits, then that display element was assigned the boundary value (this occurred less than five percent of the time). Therefore, the display values possible for both the Signal and Noise condition could span the range of allowable values, but the average value for the Signal condition was 2.4 and the average value of the Noise condition was 1.6.

Display durations of 1000 ms, 500 ms, 233 ms, and 117 ms were used for analog gauges. Display durations of 1000, 500, and 233 ms were used in the numerical display conditions. The 117 ms duration was omitted due to chance performance levels. The purpose of progressively shortening the display duration was two-fold: (1) to assess the differential effects of display duration on display type, format and number; (2) to determine the display durations to be used in the second phase of the experiment. Practice trials were conducted at the 1000 ms duration. Two hundred practice trials (two blocks) of each possible display were administered. Because practice effects were still clearly evident, the first two blocks of 500 ms trials were considered as practice trials as well and were not considered in the data analysis. Due to time constraints, subjects received practice in the single element display condition, but the single element case was eliminated in the testing phase.

Testing was conducted in 100 trial blocks. Display format, number of display elements, and display duration was constant for each block.

Eight blocks were completed during each two hour session. The 500 ms condition was tested first. All testing was completed at one duration before shortening display duration. Order of display presentation was counterbalanced such that each display _format_ appeared twice each day. Each possible Number of Display Elements, _n_, also appeared twice each day. Subjects received two blocks (200 trials) of each possible display combination, except at the 117 ms duration, when only 100 trials were administered.

Phase Two. The procedure, counterbalancing, trial sequence, stimuli, and instructions given to the subjects in the second phase of the experiment were identical to those used in Phase One. However, the displays were always comprised of nine (9) elements. The display durations used insured that all information available from the nine elements could not be fully utilized. These asymptotic performance display durations were determined for each display format from data collected in the first phase of the experiment. A display duration of 117 ms for the both analog formats and a display duration of 233 ms for both numeric formats was used. Twelve blocks (1200 trials) for each display format were conducted to facilitate accurate parameter estimation.

Experiment 2

Subjects

Two male and two female subjects were used. All subjects had normal or corrected-to-normal vision. Compensation was identical to

that received by subjects in Experiment 1. Subjects were tested in daily two-hour blocks for six weeks.

Apparatus and Procedures

This was an exact replication of the first experiment except for the following. Display density was changed such that equal visual eccentricity was provided in both the linear and square matrix array display formats. In the linear format with more than one display, the two outer positions were always filled. For display sizes greater than two, the displays were equally spaced between the two outer display elements. For display sizes greater than two in the square matrix array format, the four corner positions of the array were always filled and the remaining elements were equally spaced between the four corner elements. This density will be hereafter referred to as the Total Visual Angle (TVA) condition.

Subjects were seated 41.6 inches away from the monitors. This resulted in a maximum visual angle of: Horizontal = 12.86 degrees; and Vertical = 8.23 degrees.

All practice trials were conducted at 1000 ms duration. A total of 400 practice trials of each possible condition were administered. Data for the single element condition was collected in this experiment at all durations. During Phase 2 of this experiment, data over 20 blocks or 2000 trials per display type was collected in order to allow more precise parameter estimation.

RESULTS

Performance was greatly affected by display type and arrangement, the number of display elements, and display duration. The effects of display type, number of display elements and display duration for each display density condition were highly significant. Table A1 contains a synopsis of the ANOVA results.

The standard error of the data points for each condition from an individual subject was typically 0.2 d' units or less. In general, the plots of individual subject data were similar across all experimental conditions. The slopes of the functions, the level of performance on different display formats, and the number of elements at which the functions reach asymptote were similar and consistent for all subjects. Consequently, averaged subject data will be reported.

Figures 1 through 6 are plots of log d' vs. log n, showing the performance measure d' as a function of n, averaged over the four subjects in each display density condition for the 500 ms, 233 ms, and 117 ms durations, respectively. The ideal square-root-of-n function has been added to each figure. Several aspects of the data are evident from these figures. In Figures 1 and 2, for both display density conditions, the data points closely match the ideal function initially, but quickly fall below ideal performance levels. The amount of deviation from the ideal increases as the number of display elements is

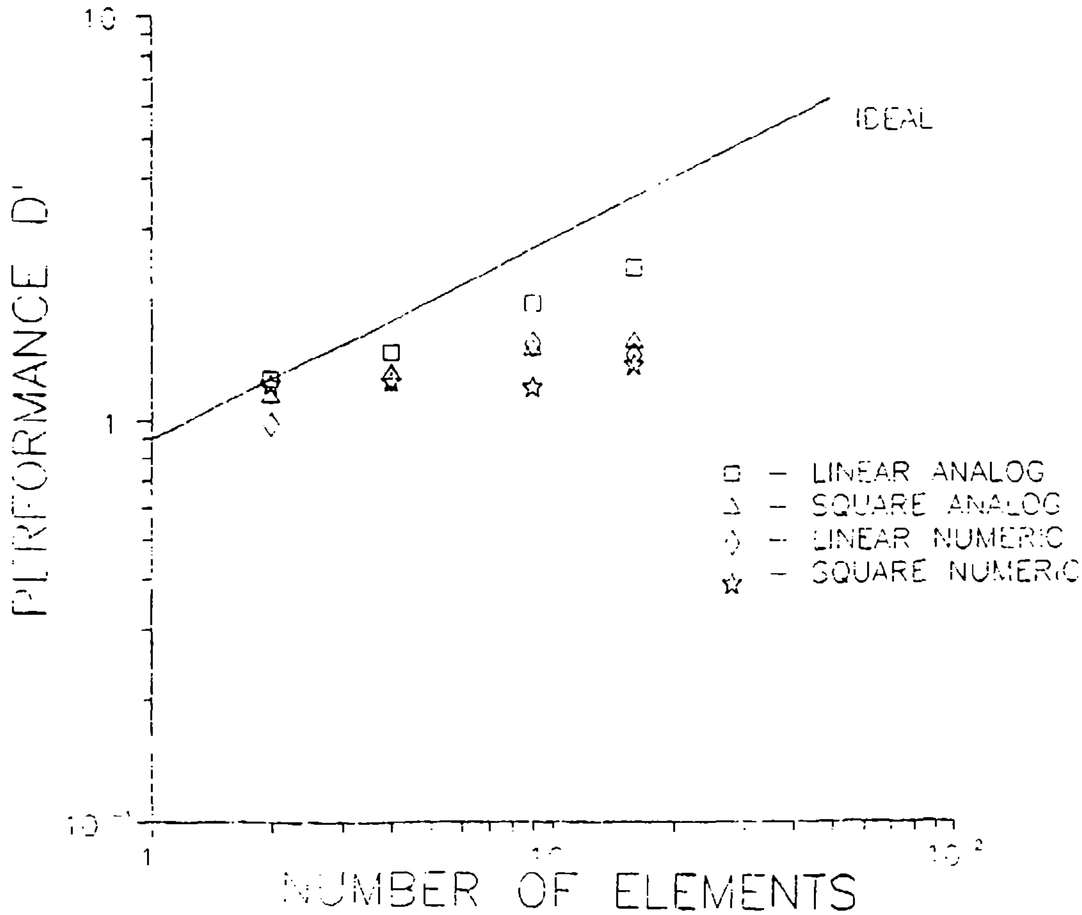

Figure 1. Average detection performance, d', as a function of the number of display elements, n, for the 500 ms conditions (Uniform Density).

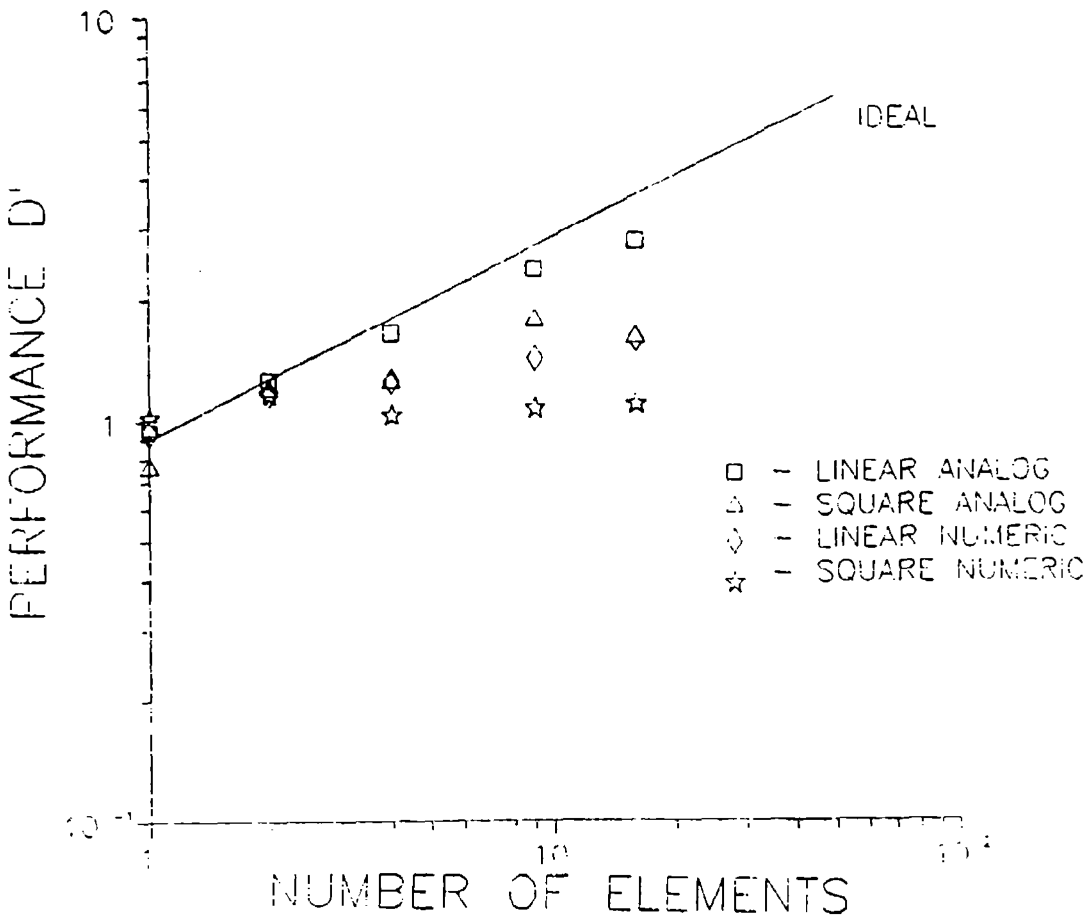

Figure 2. Average detection performance, d', as a function of the number of display elements, n, for the 500 ms conditions (Constant Total Visual Angle).

increased. Greater deviation is evident in the numeric formats, with the square numeric format exhibiting the greatest deviation. These deviations are more evident in the Total Visual Angle (TVA) display density condition. For both display density conditions, the linear analog format shows the least deviation from ideal performance.

Differences in performance across display formats are better illustrated in Figures 3 and 4, which show the data for the 233 ms duration. Performance in the linear analog format in both density conditions continues to increase as the number of display elements is increased. Performance in the square analog condition appears to have reached asymptote by 9 elements. Performance in the two numeric formats appears to asymptote around 2 elements.

In Figures 5 and 6, which show the data for the 117 ms duration, performance in the square analog condition now appears to asymptote by 4 elements. Performance in the linear analog condition, while deviating more from the ideal, still continues to increase as the number of display elements is increased. Except for the linear analog condition (for both display densities), the number of elements at asymptote decreases in an orderly and consistent fashion as the display duration is shortened.

The Partitioned Variance Model was evaluated by making least-squares estimations of the peripheral and central internal variance parameters for the averaged subject data for all display conditions at all display durations. In every case, the resultant estimate for the V_p term was a negative number. We also fit the data to a model with a

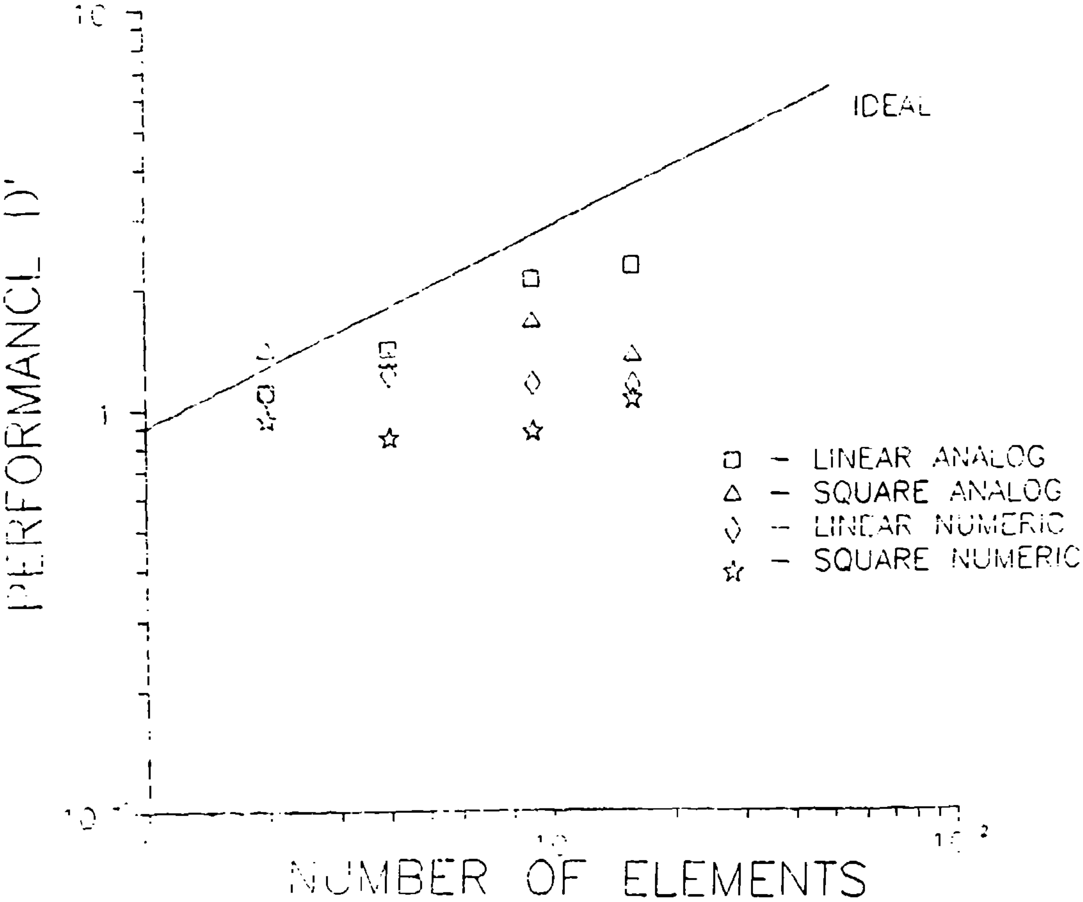

Figure 3. Average detection performance, d', as a function of the number of display elements, n, for the [illegible]3 ms conditions (Uniform Density).

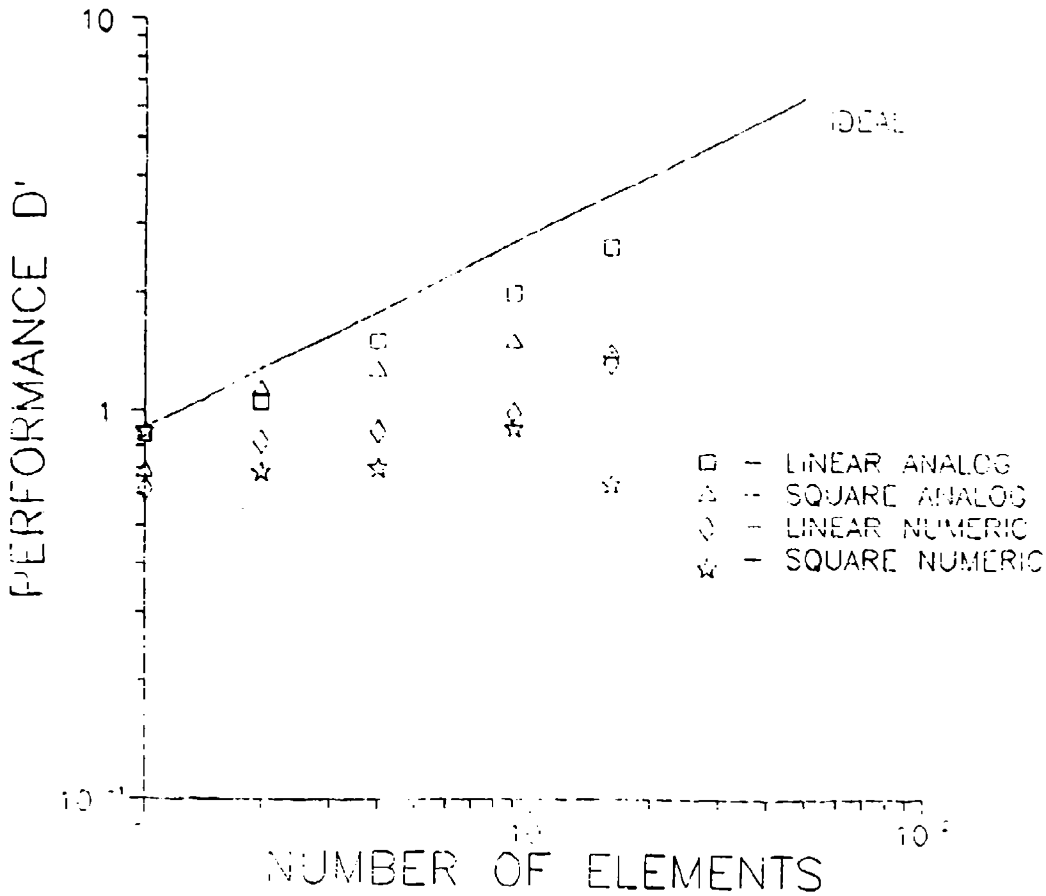

Figure 8. Average detection performance, d', as a function of the number of display elements, n, for the 233 ms conditions (Constant Total Visual Angle).

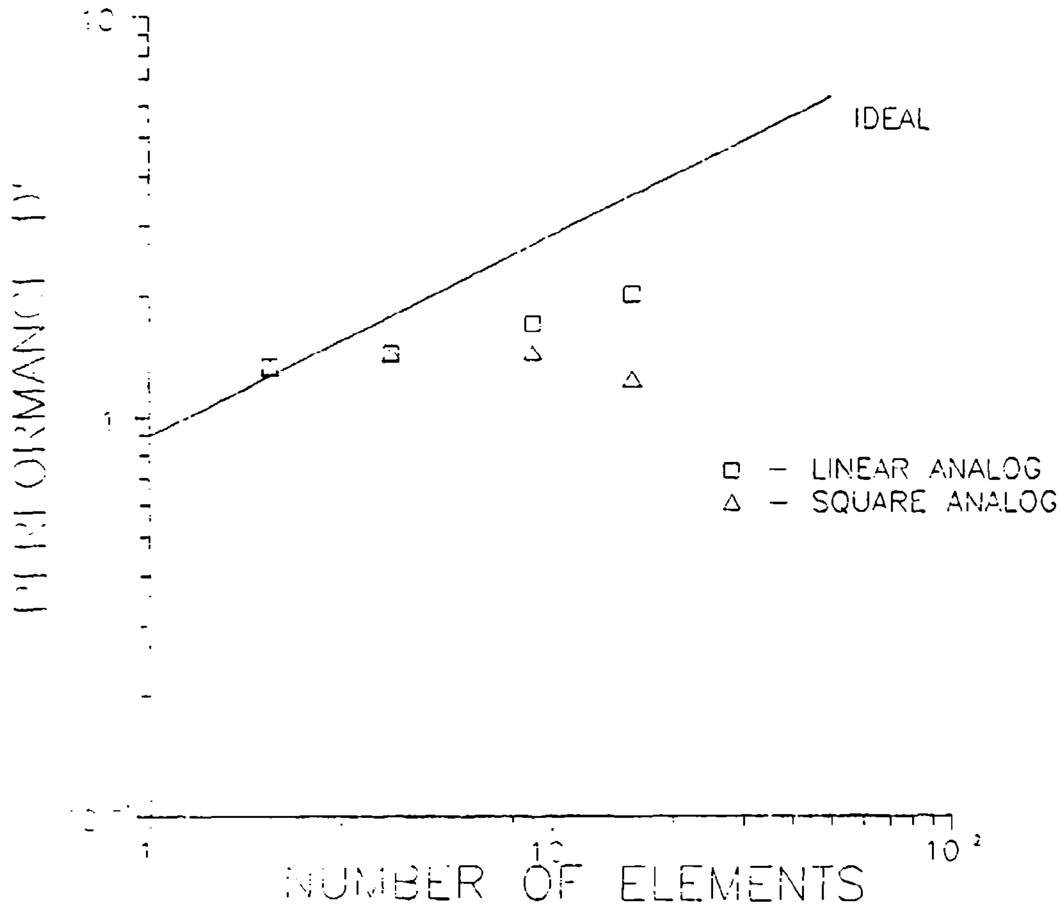

Figure 5. Average detection performance, d', as a function of the number of display elements, n, for the 117 ms conditions (Uniform Density).

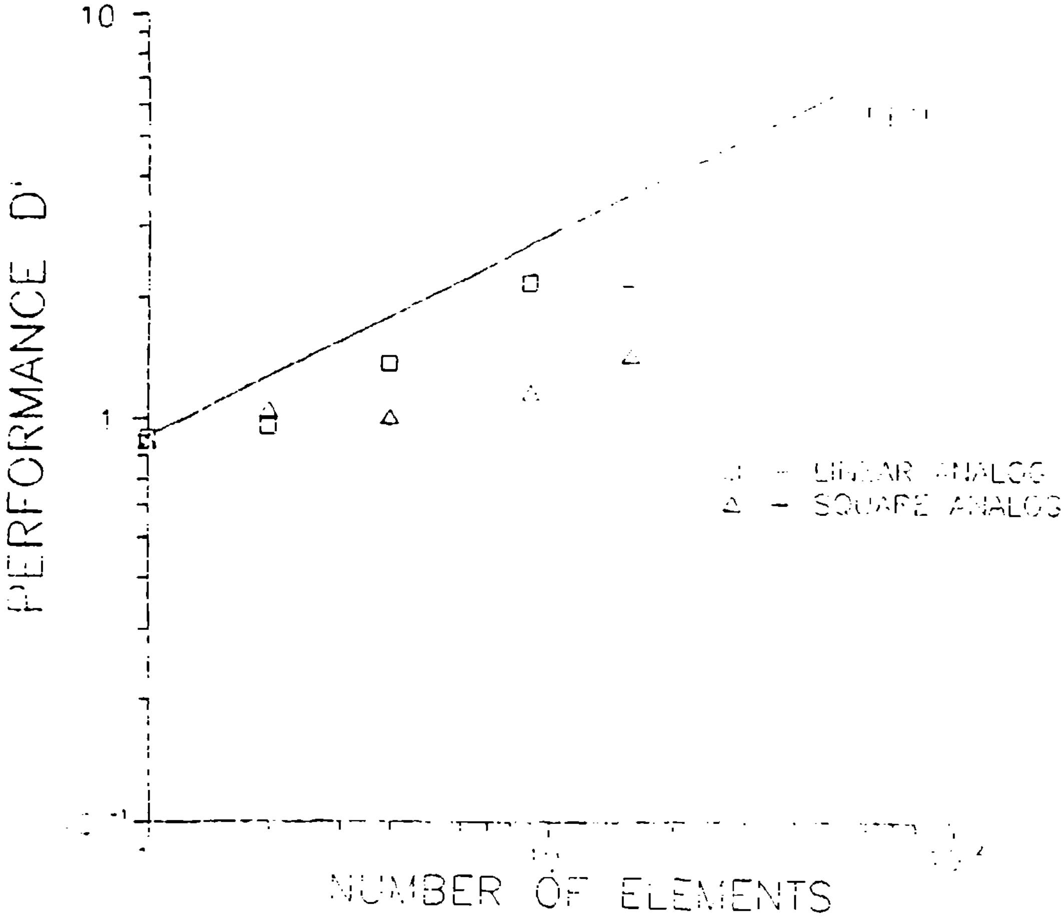

Figure 6. Average detection performance, d′, as a function of the number of display elements, n, for the 117 ms conditions (Constant Total Visual Angle).

single fixed internal variance parameter. The resultant performance equation can be written:

$$d'_{(n)} = (m2 - m1) / (v_{ext}/n + v_f)^{1/2} \qquad [4]$$

where V_f represents a fixed variance term that is not dependent on the number of display elements. This equation provided a reasonable fit to the data for all of the possible display conditions at all display durations, although the fits were better for some display conditions than others. Two representative examples are depicted in Figure 7. It shows the plot of d' vs n for averaged subject data. The prediction of the fixed variance equation and the resultant V_f value for the two conditions are shown. V_f values and the maximum and average absolute deviations of the fit to the data for each condition are listed in Table 1.

The analytical procedure from Robinson et al. (1986) was used to estimate whether some display elements weigh more heavily in a subject's decision than information from other elements in the display. The basic reasoning behind the analysis is as follows: Recall that on each trial, the subject views the display and responds "Signal" or "Noise." For a single element display, the response is made by comparing the displayed value, x_j (where x is of the 41 possible display values; 0.0 to 4.0 in 0.1 increments), against some decision criterion, C. If $x_j < C$, then the response is "noise." If $x_j > C$, then the response is "signal." For an ideal observer, having no internal noise, the probability function p("signal"/x_j), relating the probability of responding "signal" given display value x_j, would be a

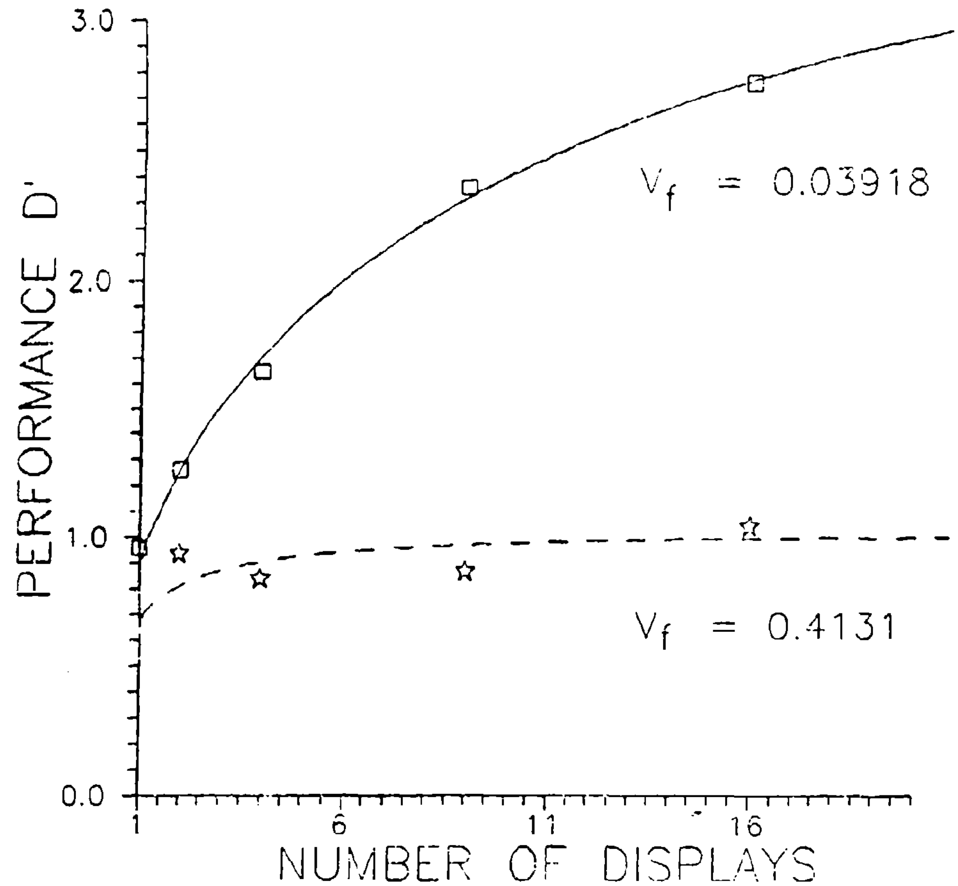

Figure 7. Two examples of the Fixed Variance Model (see text) fits to averaged subject data. The solid line depicts the best fit to the Linear Analog (Constant Total Visual Angle) condition data (square symbols) at a display duration of 500 ms. The dashed line depicts the best fit to the Square Numeric (Uniform Density) condition data (star symbols) at a display duration of 233 ms.

Table 1. Fixed Va. iance Model parameter estimates f r V_f, total internal variance, for each display condition and d lay duration expressed in display units. The maximum and average absolute deviations of the model fit to the data for each condition are included.

Duration (in ms)	Display Condition	V_f	Max Dev	Average Dev
500	Linear Analog UD*	0.066	0.098	0.071
500	Linear Analog TVA*	0.039	0.110	0.063
500	Square Analog UD	0.228	0.150	0.109
500	Square Analog TVA	0.179	0.235	0.108
500	Linear Numeric UD	0.223	0.123	0.073
500	Linear Numeric TVA	0.258	0.173	0.114
500	Square Numeric UD	0.343	0.308	0.148
500	Square Numeric TVA	0.493	0.319	0.144
233	Linear Analog UD	0.056	0.132	0.082
233	Linear Analog TVA	0.053	0.143	0.102
233	Square Analog UD	0.200	0.373	0.210
233	Square Analog TVA	0.224	0.124	0.091
233	Linear Numeric UD	0.316	0.163	0.117
233	Linear Numeric TVA	0.464	0.207	0.110
233	Square Numeric UD	0.594	0.136	0.086
233	Square Numeric TVA	0.925	0.271	0.120
117	Linear Analog UD	0.145	0.250	0.149
117	Linear Analog TVA	0.055	0.277	0.161
117	Square Analog UD	0.342	0.399	0.253
117	Square Analog TVA	0.341	0.143	0.105

*UD - Uniform Density
*TVA - Constant Total Visual Angle

step function. That is, when $x_j < C$, the ideal always responds "noise." When $x_j > C$, the ideal always responds "signal." Deviations from this step function can be attributed to internal noise. In general, this function will have the form of a cumulative normal probability distribution, where the magnitude of the internal noise is related to the standard deviation of the function. The standard deviation of this function can be obtained by a least squares fit of the data to a cumulative normal probability function. If the slope of the resulting ogive were small, we could conclude that the subject's response appeared to be nearly independent of the value of the display element. Conversely, if the slope of the ogive were large, we could conclude that the subject's response was highly dependent on the value of the display element. When applied to each element of a multi-element display, this analysis yields estimates of the slope of the ogive for each display position; these slopes provide an index of the relative contribution of each element to the decision statistic (Berg, 1987 has a more complete discussion of the assumptions and limitations of the method).

In order to make the slope estimates, the 41 possible display values for each display element position were assigned to 13 bins. The width of each bin, except for the first and last bins which were 0.3 display units wide, was 0.2 display units. The SAS Institute, Inc. Categorical Data Modeling (CATMOD) procedure was used to estimate the slope parameter for each element. CATMOD uses a weighted-least-squares method to minimize the weighted residual sum of squares when estimating parameters. Figure 8 depicts theoretical probability functions

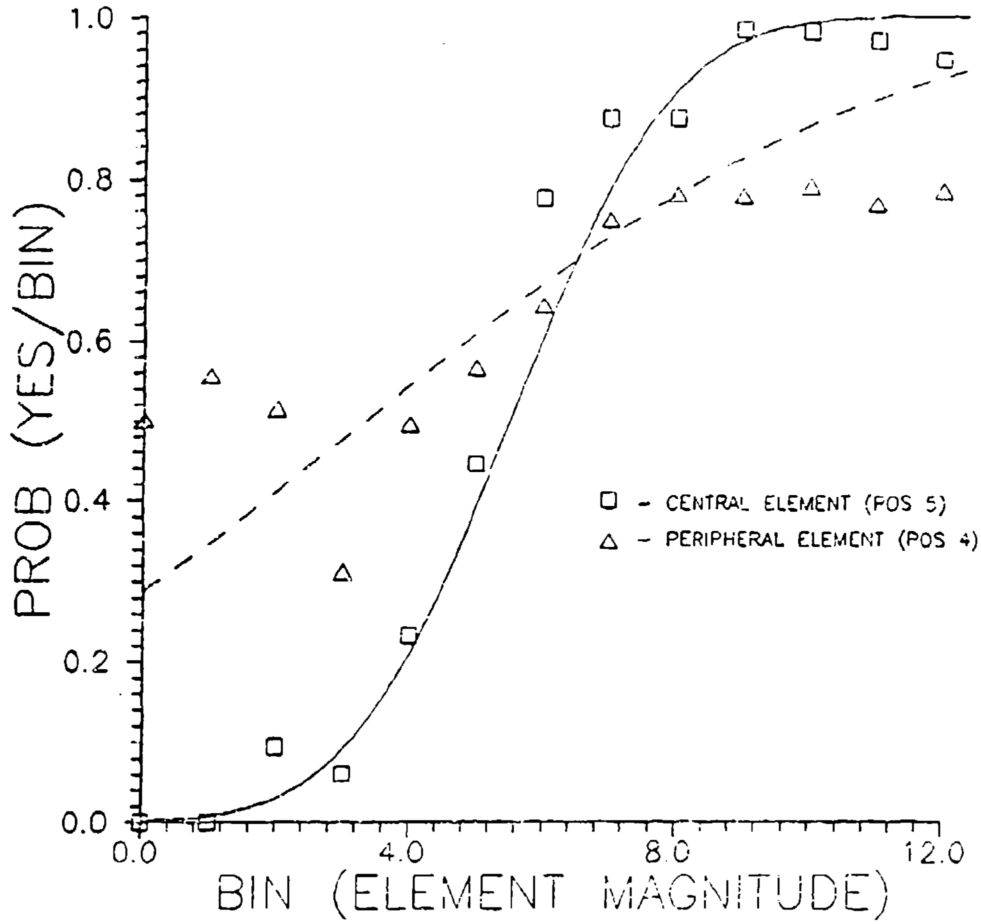

Figure 8. Theoretical and obtained response probability functions (see text) for Subject TZ, Square Numeric (Constant Total Visual Angle) condition, for central element 5 (square symbols) and peripheral element 4 (diamond symbols).

generated from the parameter estimates calculated for element position Four and Five for observer TZ in the Square Numeric (Constant Total Visual Angle) condition at a display duration of 233 ms. Comparison of these theoretical curves to actual data points indicate reasonable fits to the data when the response function was both highly dependent on display value (element five, labeled "central element") or when the response appeared nearly independent of the display values (element four, labeled "peripheral element").

Figures 9 through 16 depict averaged slope estimates obtained for each display position for each experimental condition. The ordinate of each figure is the slope associated with the display element at each display position (in units of 1/bins). Differences between display types are readily apparent. Results for the Linear Numeric formats are shown in Figures 9 and 10. The major contribution to the subjects' responses came from the three elements in the center of the display. The three elements at the extreme right and extreme left of the display had relatively little influence. Planned comparisons of slope values utilizing the CATMOD CONTRAST procedure indicated that the slope values of the center three elements was significantly higher than the slope values of the rest of the display elements for each subject ($p < .001$). Figures 11 and 12 depict the Linear Analog conditions. Now it appears that information from many of the elements has an influence on the response. While the three center positions still had a significantly greater slope value for six of the 8 subjects, the difference was not as great as in the Linear Numeric condition (typically $p < .05$). Two subjects showed no significant differences.

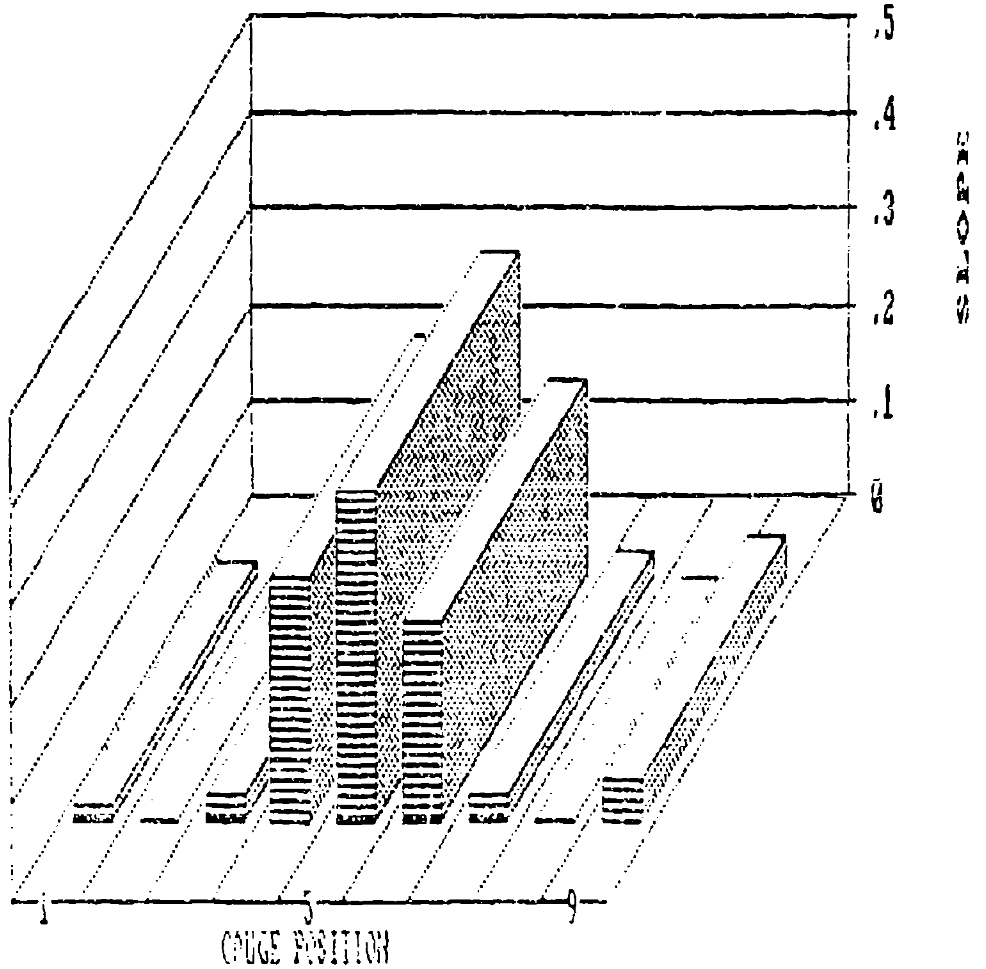

Figure 9. Average estimated slopes of the response probability function (see text) for each element spatial position in the Linear Numeric (Uniform Density) condition at a display duration of 233 ms.

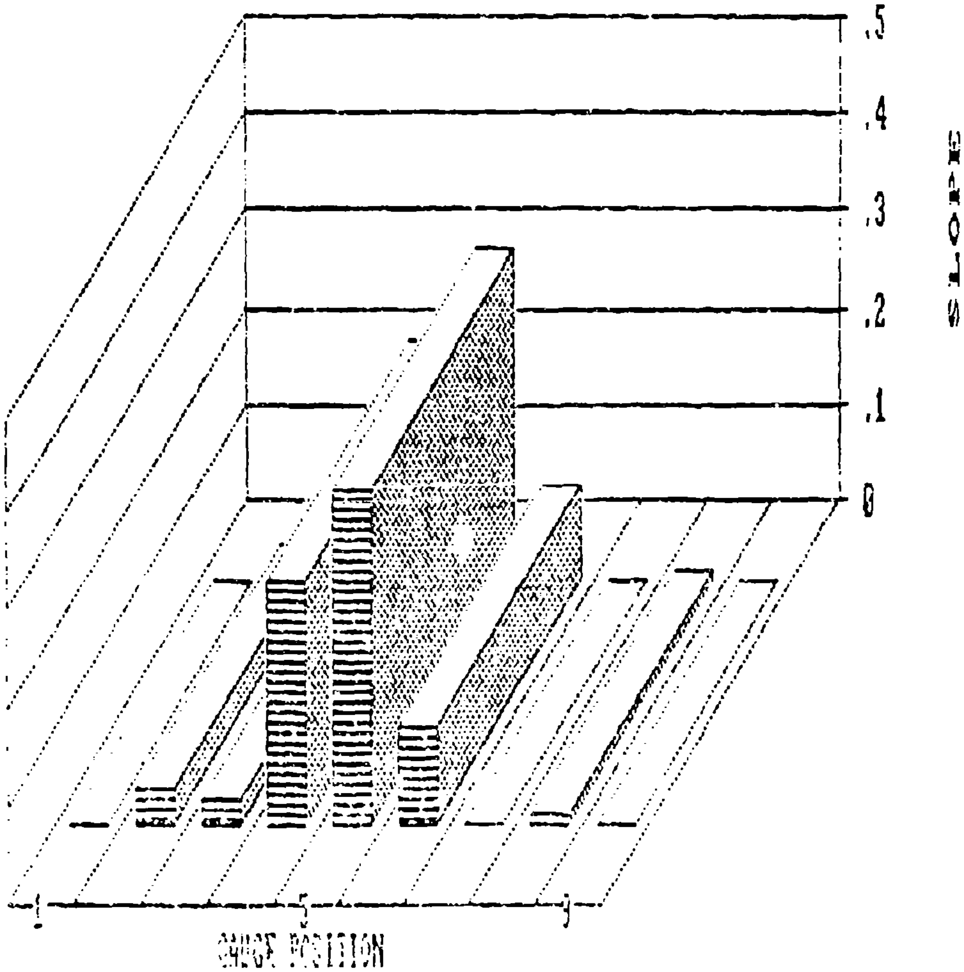

Figure 10. Average estimated slopes of the response probability function (see text) for each element spatial position in the Linear Numeric (Constant Total Visual Angle) condition at a display duration of 233 ms.

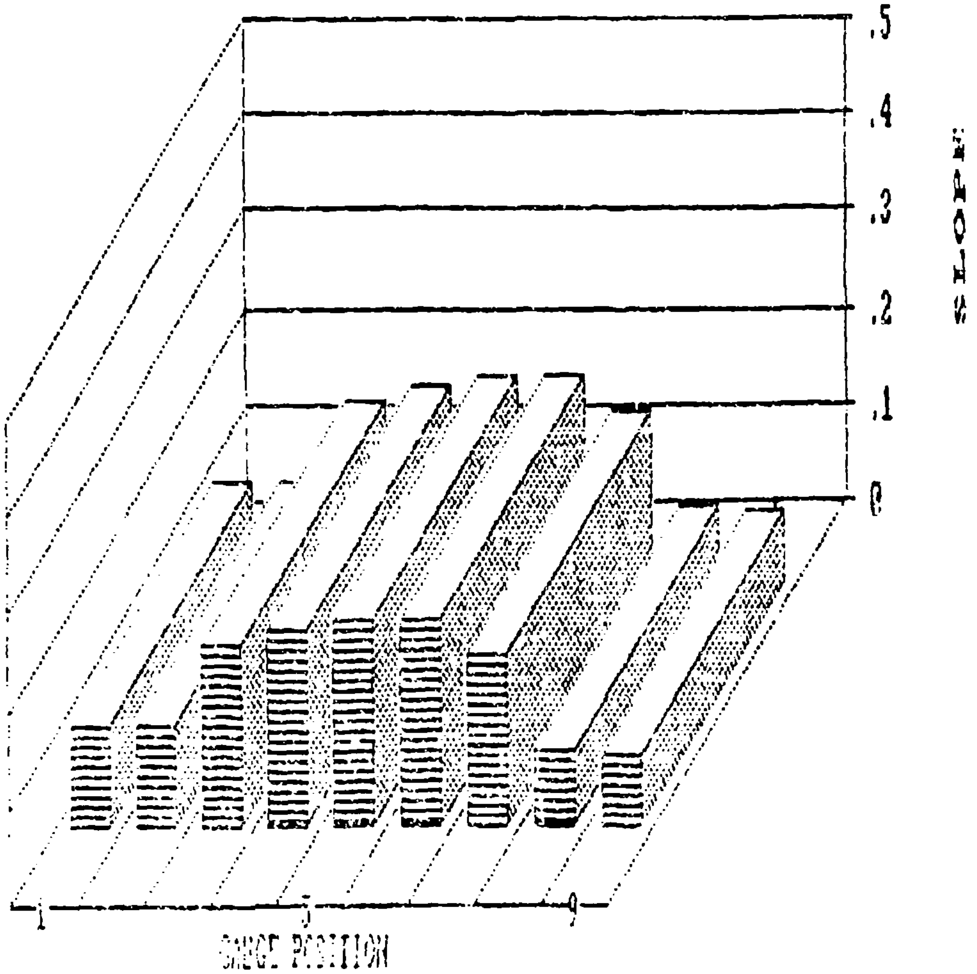

Figure 11. Average estimated slopes of the response probability function (see text) for each element spatial position in the Linear Analog (Uniform Density) condition at a display duration of 117 ms.

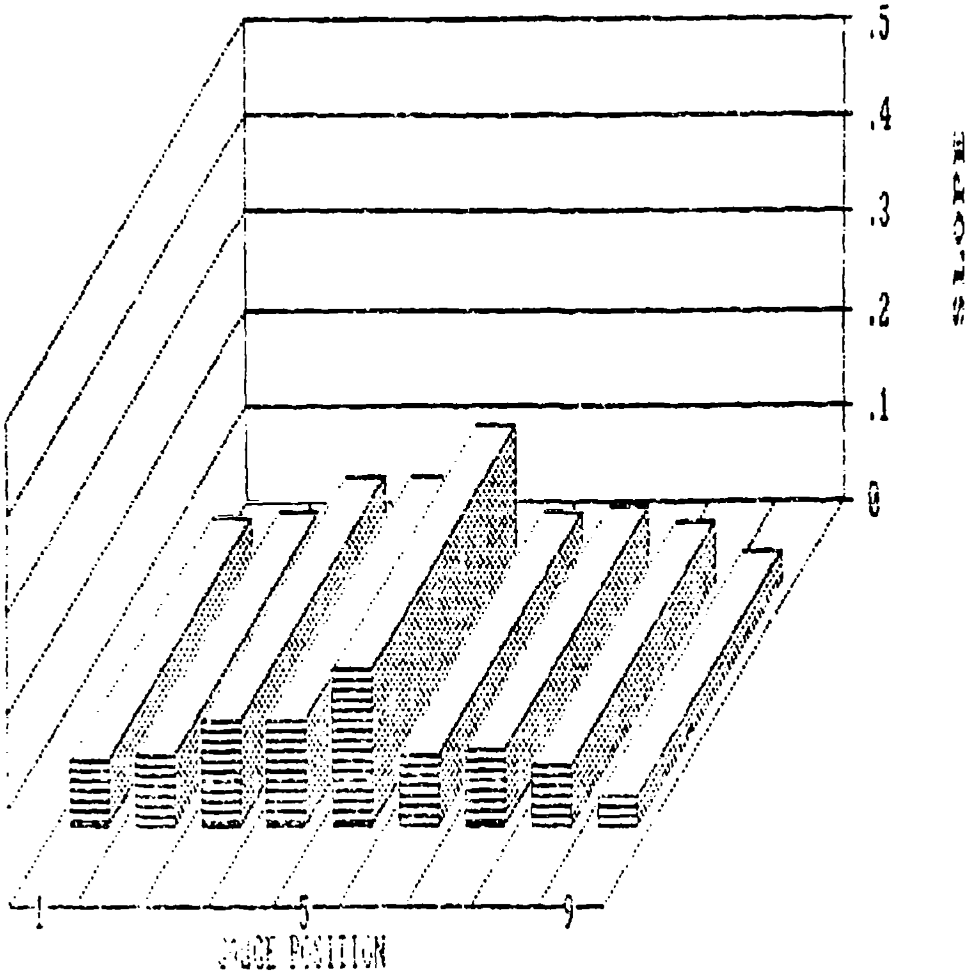

Figure 12. Average estimated slopes of the response probability function (see text) for each element spatial position in the Linear Analog (Constant Total Visual Angle) condition at a display duration of 117 ms.

Figures 13 and 14 depict results from the Square Analog conditions and Figures 15 and 16 show results from the Square Numeric conditions. A picture similar to that in the linear arrangement emerges: information is accrued from many display elements in the analog displays, while from just one element in the numeric displays. A planned comparison analysis showed that for the square numeric conditions, the display position with the highest slope (usually the center position) had a significantly greater slope value than any of the other elements ($p < .001$ for seven of the 8 subjects). This was not the case for the square analog condition. When the two greatest slope values were compared, no significant differences were found for five of the 8 subjects. For the three remaining subjects, the differences were not as great as in the Square Numeric condition ($p < .05$, $p < .02$, $p < .004$).

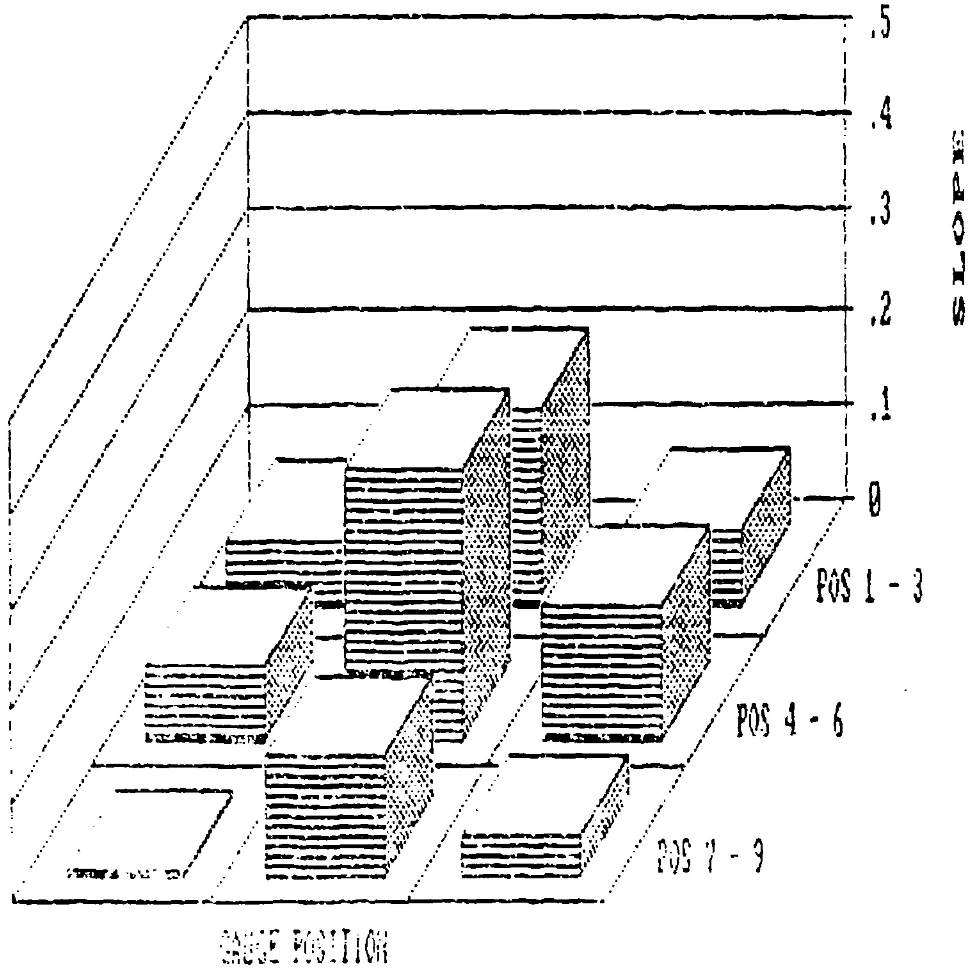

Figure 13. Average estimated slopes of the response probability function (see text) for each element spatial position in the Square Analog (Uniform Density) condition at a display duration of 117 ms.

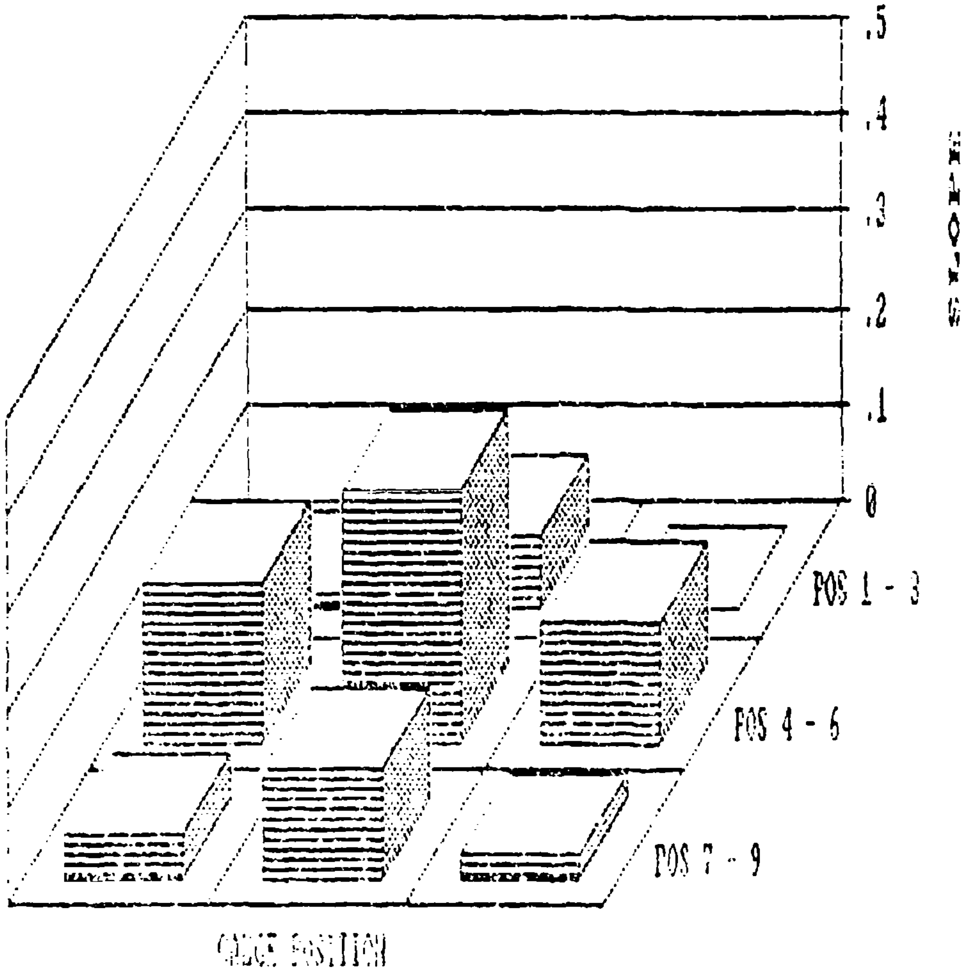

Figure 14. Average estimated slopes of the response probability function (see text) for each element spatial position in the Square Analog (Constant Total Visual Angle) condition at a display duration of 117 ms.

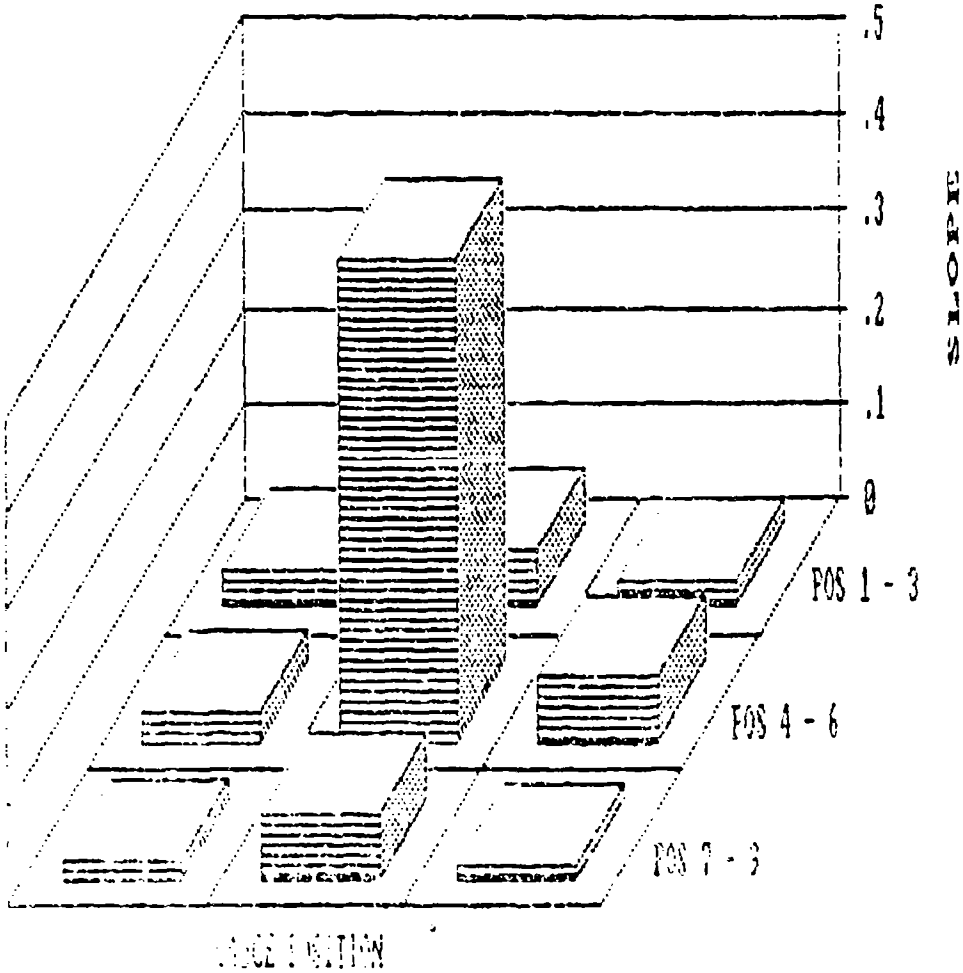

Figure 15. Average estimated slopes of the response probability function (see text) for each element spatial position in the Square Numeric (Uniform Density) condition at a display duration of 233 ms.

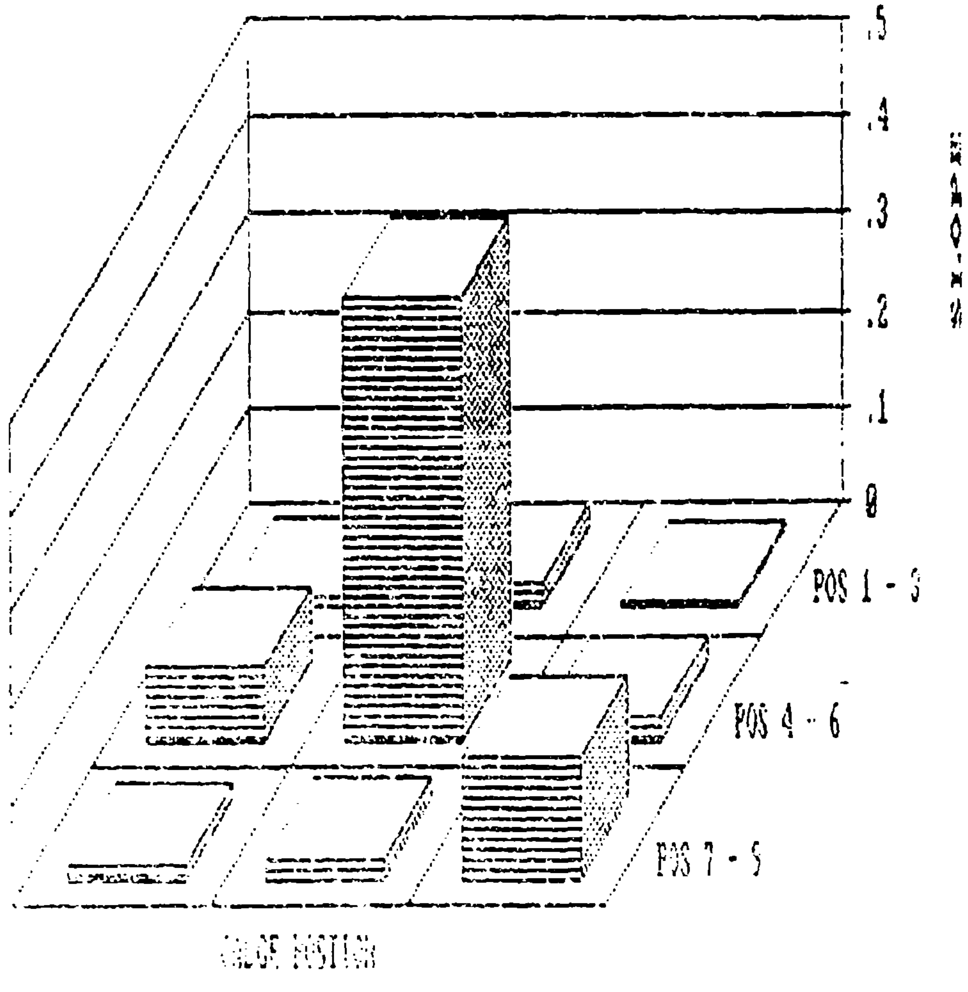

Figure 16. Average estimated slopes of the response probability function (see text) for each element spatial position in the Square Numeric (Constant Total Visual Angle) condition at a display duration of 233 ms.

DISCUSSION

Display duration and display format greatly influenced subject performance. Performance increased at close to the ideal rate up to different asymptotic levels that depended on display format and display duration. Using some display formats, subjects were able to aggregate information from as many as sixteen display elements. In many longer duration and small number of element conditions, subjects were able to use all the information available to make their decision. In this task, subjects can more efficiently process information from analog display elements. When the display duration is short, the amount of information that may be obtained from a display is highly dependent on the arrangement of the display elements.

It is not surprising that subjects were better able to aggregate information from display elements that were arranged in a linear format compared to a square matrix array format. Obviously, only one row of displays had to be scanned in the linear format, compared to as many as four rows in the square matrix array format. The fact that the same number of elements arranged in a square matrix array covered more display area than the same number of elements presented linearly may have also contributed to the disparity in performance levels (see Tullis (1983) for a discussion of "local density" and its effects on performance).

Comparisons of the present results to those obtained in the Sorkin and Weldon study demonstrate that the order of stimulus presentation has little effect on performance. The overall performance levels of the two studies are very similar. The effects of display type, arrangement, and display duration on detection performance are also similar and consistent. In psychophysical experiments such as this, the order of stimulus presentation appears to have little or no influence on performance.

An interesting result of the display format comparison is that performance approximates the ideal up to different asymptotic levels for all display conditions. Shortening the display duration appears to effect performance by reducing the number of display elements that can be processed by the subject, not by reducing the amount of information in the displays. This hypothesis is supported by data obtained using the Robinson et al. (1986) analytical procedure. When processing information from digital displays, only a few central elements seem to influence a subject's decision. The subject's response is highly dependent on the values presented at these central elements positions. For the remaining spatial positions, a subject's response is essentially independent of the value presented. These "peripheral" elements had little or no influence on a subject's response.

When information is processed from analog displays, a greater number of elements seem to have a bearing on a subject's decision. Fewer elements can be considered to be "peripheral" in an information processing sense. For a given display duration, a greater number of analog display elements can be sufficiently processed to influence the

subject's decision. Being able to consider more elements in a given amount of time increases the probability of a correct response. This may partially account for higher performance with analog displays.

Further support for this hypothesis could be gained by showing that the number of influential elements decreases as the display duration is decreased. Because this study used a single display duration for each display type when evaluating the relative influence of each spatial position, additional research is needed to test this hypothesis.

The Robinson et al. (1986) analytical procedure proved to be a useful method for assessing the relative importance of specific spatial elements of a multi-element display. It showed clearly that some elements in a multi-element display have more influence on a decision than other display elements. The relative influence of a particular spatial element and the total number of elements that influence the decision also appears to be highly dependent on display type and arrangement.

Why is performance in this task better with analog displays than with digital displays? It may be that the digital displays require additional encoding or decoding to transform them into an internally usable form. The analog displays may be processed faster because they impose easier or fewer conversion operations. A study by Hanson et al. (1981) supports such an hypothesis. In a study comparing check reading performance with numerical and analog displays, digital displays yielded significantly longer detection times. Performance with digital displays was also more sensitive to several other experimental

manipulations including the number of display elements to be monitored. These results are consistent with those of the present experiment. In terms of the visual processing model proposed by Treisman (1986), it may be that the descriptions stored in the "recognition network" must be accessed to allow for full processing of digital displays. This "recognition network" specifies the critical aspects of familiar perceptual objects, allowing access to their names, their likely behavior and their current significance. It is possible that digital elements must be more fully "named" or decoded prior to being transformed into a more internally usable form.

Robinson, Grantham, and Berg's (1986) Partitioned Variance Model was not supported by least-squares model fits to the averaged data. In every case, the resultant estimate for the V_p term was a negative number. This indicates that performance with a small number of elements increases more rapidly than the ideal rate. This is consistent with a model discussed by Nolte (1967). This model suggests that an observer who is initially uncertain about an aspect of the signal (such as the average gauge value given signal), may exhibit faster than $n^{1/2}$ improvement in performance. In the present experiment the observers were highly practiced and explicitly informed about the value of the mean gauge readings. Thus, it seems unlikely that uncertainty could account for the present result. Reasonable fits were obtained to a simpler model having a single, fixed internal variance term. The magnitude of this fixed variance term depended on the display format and arrangement, but display density (element spacing) had little effect on its value. The value of V_f is a good summary

indicator of performance, since it is the parameter that determines the function's asymptotic value. It is likely that there are internal noise sources of both peripheral and central origin. Much of the peripheral noise associated with individual elements may not be independent across elements. For example, if internal noise were correlated across the display elements, then performance would not continue to increase as the number of display elements was increased. The assumption of correlated internal noise in the visual system is not unreasonable, especially if one considers the processing of closely spaced elements. The processes involved in transforming display element information into internally usable form may be susceptible to interference from internal noise processes that are correlated across adjacent spatial channels. Increasing the similarity or proximity of the display elements or decreasing the display duration could increase the effective correlation of the noise.

This study has shown that a TSD approach provides a useful framework for investigating visual display information processing. Detection theory methodology allows specification of important parameters of visual display processing and facilitates the comparison of different display types. The analytical procedure developed by Robinson et al. (1986) allows for useful comparisons to be made between different types and arrangements of visual displays. The method can be used to assess the relative importance of specific spatial elements of a visual display and may provide insights into the manner by which an observer accumulates information from visual displays. It is hoped

that the knowledge gained from this approach can be used to aid the design of practical visual display systems.

REFERENCES

REFERENCES

Brunswik, E. (1956). Perception and the representative design of experiments. Berkeley: University of California Press.

Berg, Bruce G. (1987). Internal noise in auditory decision tasks. Unpublished doctoral dissertation, Indiana University.

Crawford, A. M. (1977). Interactive computer graphics for simulation displays. Proceedings of the Human Factors Society 21st Annual Meeting. 14-17.

DeSanctis, G. (1984). Computer graphics as decision aids: Directions for research. Decision Sciences, 15, 463-487.

Dudycha, L. W., & Naylor, J. C. (1966). Characteristics of the human inference process in complex choice behavior situations. Organizational Behavior and Human Performance, 1, 110-128.

Grace, G. L. (1966). Application of empirical methods to computer-based system design. Journal of Applied Psychology, 50, 442-450.

Green, D. M. & Swets, J. A. (1974). Signal Detection Theory and Psychophysics. Huntington, N.Y.: Keiger. (Original work published 1966)

Hanson, R. H., Payne, D. G., Shively, R. J., & Kantowitz, B. H. (1981). Process control simulation research in monitoring analog and digital displays. Proceedings of the Human Factors Society 25th Annual Meeting. 154-158.

Kahneman, D. (1973). Attention and effort. Englewood Cliffs, N.J.: Prentice-Hall.

Lincoln, R. S., & Cahill, H. E. (1965). Detecting out-of-tolerance conditions with meter and digital displays. Human Factors, 54-62.

MacGregor, D., & Slovic, P. (1986). Graphic representation of judgmental information. Human-Computer Interaction, 2, 179-200.

Navon, D., & Gopher, D. (1979). On the economy of the human processing system. Psychological Review, 86, 214-255.

Nawrocki, L. H. (1972). Alphanumeric versus graphic displays in a problem solving task. (Technical Research Note 227). Arlington, VA: U.S. Army Behavior and Systems Research Laboratory. (NTIS No. AD748799)

Nolte, L. W. (1967). Theory of signal detectability: Adaptive optimum receiver design. Journal of the Acoustical Society of America, 42(4), 773-777.

Norman, D., & Bobrow, D. (1975). On data limited and resource limited processing. Journal of Cognitive Psychology, 73, 44-60.

Remington, R., & Williams, D. (1986). On the selection and evaluation of visual display symbology: Factors influencing search and identification times. Human Factors, 28(4), 407-420.

Robinson, D. E., Grantham, W. & Berg, B. G. (1986). A partitioned variance model for multiple observations. Paper presented at the Nineteenth Annual Meeting of the Society for Mathematical Psychology, Cambridge, MA.

Sorkin, R. D., Robinson, D. E., & Berg, B. G. (1987). A detection theory method for the analysis of visual and auditory displays. Proceedings of the Human Factors Society 31st Annual Meeting, 2, 1184-1187.

Sorkin, R. D., & Weldon, M. (1987). A detection theory method for evaluating visual displays. Unpublished manuscript.

Spragg, S. D., & Rock, M. L. (1952). Dial reading performance as a function of brightness. Journal of Applied Psychology, 36, 128-137.

Stock, D., & Watson, C. J. (1984). Human judgment accuracy, multidimensional graphics, and humans versus models. Journal of Accounting Research, 22(1), 192-206.

Swets, J. A. (1984). Mathematical models of attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of attention (pp. 183-242). Orlando, FL: Academic Press.

Treisman, A. (1986). Features and objects in visual processing. Scientific American, 225(5), 114B-125.

Tullis, T. S. (1981). An evaluation of alphanumeric, graphic, and color information displays. Human Factors, 23(5), 541-550.

Tullis, T. S. (1983). The formatting of alphanumeric displays: A review and analysis. Human Factors, 25(6), 657-682.

Vincino, F. L., & Ringel, S. (1966). Decision-making with updated graphic vs. alphanumeric information. (Technical Research Note 178). Washington, D.C.: Army Personnel Research Office. (NTIS No. AD647623)

Wickens, C. D. (1984a). Engineering psychology and human performance. Columbus, OH: Merrill.

Wickens, C. D. (1984b). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of attention (pp. 63-102). Orlando, FL: Academic Press.

APPENDIX

Table A1. Synopsis of ANOVA results for both the Uniform Density and Constant Total Visual Angle conditions. The error term used for the hypotheses tests and P-values are listed.

Uniform Density Condition

Source	DF	Type III SS	F Value	P ≥ F
Duration (error term - subject*duration)	2	1.137	16.18	0.0035
Type (error term - subject*type)	3	14.699	136.19	0.0001
Number (error term - subject*number)	3	3.980	7.20	0.0091

Constant Total Visual Angle Condition

Source	DF	Type III SS	F Value	P ≥ F
Duration (error term - subject*duration)	2	7.299	34.73	0.0005
Type (error term - subject*type)	3	25.348	49.26	0.0001
Number (error term - subject*number)	3	24.941	34.77	0.0001